Article

A Fuzzy Synthetic Evaluation Approach to Assess Usefulness of Tourism Reviews by Considering Bias Identified in Sentiments and Articulacy

Dimitrios K. Kardaras [1,*] **, Christos Troussas** [2] **, Stavroula G. Barbounaki** [3] **, Panagiota Tselenti** [2] **and Konstantinos Armyras** [1]

1   Business Informatics Lab, Department of Business Administration, Athens University of Economics and Business, 10434 Athens, Greece; armyras@aueb.gr
2   Department of Informatics and Computer Engineering, University of West Attica, 12243 Egaleo, Greece; ctrouss@uniwa.gr (C.T.); nota@uniwa.gr (P.T.)
3   Department of Midwifery, University of West Attica, 12243 Egaleo, Greece; sbarbounaki@uniwa.gr
*   Correspondence: dkkardaras@yahoo.co.uk or kardaras@aueb.gr

Abstract: Assessing the usefulness of reviews has been the aim of several research studies. However, results regarding the significance of usefulness determinants are often contradictory, thus decreasing the accuracy of reviews' helpfulness estimation. Also, bias in user reviews attributed to differences, e.g., in gender, nationality, etc., may result in misleading judgments, thus diminishing reviews' usefulness. Research is needed for sentiment analysis algorithms that incorporate bias embedded in reviews, thus improving their usefulness, readability, credibility, etc. This study utilizes fuzzy relations and fuzzy synthetic evaluation (FSE) in order to calculate reviews' usefulness by incorporating users' biases as expressed in terms of reviews' articulacy and sentiment polarity. It selected and analyzed 95,678 hotel user reviews from Tripadvisor, written by users from five specific nationalities. The findings indicate that there are differences among nationalities in terms of the articulacy and sentiment of their reviews. The British are most consistent in their judgments expressed in titles and the main body of reviews. For the British and the Greeks, review titles suffice to convey any negative sentiments. The Dutch use fewer words in their reviews than the other nationalities. This study suggests that fuzzy logic captures subjectivity which is often found in reviews, and it can be used to quantify users' behavioral differences, calculate reviews' usefulness, and provide the means for developing more accurate voting systems.

Keywords: fuzzy logic; sentiment analysis; reviews usefulness; bias; cultural differences; tourism; Tripadvisor

1. Introduction

1.1. Sentiment Analysis

In today's digital era, understanding the underlying feelings and potential biases within users' reviews has become critical, since user-generated content has significant impact over customer decisions. Sentiment analysis, also known as opinion mining, is the computational analysis of people's views, feelings, emotions, and attitudes about entities such as products, services, issues, events, ideas, and their attributes [1]. As opinions and sentiments are widely expressed on many platforms ranging from e-commerce websites to social media, the necessity for accurate sentiment analysis technologies that not only recognize emotions but also detect potential biases has become critical, thus improving reviews' usefulness.

Analyzing the pertinent literature, it can be inferred that sentiment analysis is a rich area for research across various fields, such as e-business, e-learning, marketing, social networking sites, customer feedback, political discourse, etc.

Firstly, in the area of e-business, it has been used to assess the sentiments of customers and receive their preferences on either products or services [2–6]. Indeed, online commerce platforms can have availability of a large pool of data in the form of customer feedback or reviews that can serve as a valuable source of information towards building their promotion strategy. As such, companies apply sentiment analysis techniques on this data to gain important insights on customer satisfaction and devise strategies towards improving customer experience. A recent review on sentiment analysis in e-business attests that sentiment analysis has been widely used in e-commerce using mainly machine learning algorithms.

In the context of e-learning, sentiment analysis finds application in analyzing students' feedback on learning objects, forum discussions, and course evaluations [7–11]. This process yields valuable insights into learners' emotions and sentiments, enabling the creation of personalized learning experiences. By understanding students' feelings, educators can deliver tailored learning materials, exercises, and collaborative opportunities. Moreover, sentiment analysis serves as a valuable tool for identifying areas of improvement in the e-learning environment, helping educators enhance the overall learning experience. A 2023 review [12] highlighted that sentiment analysis has been shown to be effective for educators as it helped them improve their teaching methodology and tailor their course content to students. Also, concerning learners, sentiment analysis has helped them advance their knowledge and has provided them with access to qualitative learning.

Furthermore, sentiment analysis has also been widely used by social networking sites [13–17], since understanding the opinion of people holds significant implications. Researchers have developed domain-specific sentiment analysis techniques to tackle the unique challenges posed by noisy social media data. A 2023 review study [18] attested that sentiment analysis can have great implications in this field, while the techniques used are mainly neural networks and Support Vector Machines (SVMs).

Also, sentiment analysis finds application in political discourse analysis [19–23], where analyzing public sentiment towards political candidates, policies, and events can inform campaign strategies and policy decisions. By analyzing sentiments expressed in political tweets, news articles, and online forums, valuable insights into public perceptions can be gained.

As shown in the aforementioned studies, two prevalent approaches have been used for sentiment analysis. The first approach comprises lexicon-based techniques. Lexicon-based methods assign positive, negative, or neutral sentiment scores to individual words and then aggregate the scores to determine the overall sentiment of a text. While simple and easy to implement, lexicon-based methods suffer from limited context awareness and may struggle with sarcasm, idioms, or language nuances.

Machine learning approaches in sentiment analysis can perform better in situations where lexicon-based techniques may present obstacles. Researchers have resorted to machine learning techniques, such as supervised learning and deep learning, to categorize text sentiments automatically. Supervised learning entails training models on labelled datasets, where each text is assigned a sentiment label such as positive, negative, or neutral. These models learn to recognize patterns in the data that can be applied to new situations with similar sentiments. Among the popular supervised learning algorithms used in sentiment analysis are Support Vector Machines, Naive Bayes (NB), and Random Forests. These algorithms are adept at classifying sentiments based on the patterns they learn from the labelled training data.

Deep learning approaches, including Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), have shown important effectiveness in sentiment analysis applications. RNNs perform well for sequential data and can capture long-term dependencies in text. CNNs are well-suited in learning local patterns and features from text. Moreover, the advent of pre-trained language models, such as BERT and GPT, has significantly advanced sentiment analysis, as these models can have high-quality performance on specific sentiment analysis tasks.

In conclusion, a recent review by [24] highlights the extensive application of sentiment analysis, with social networks being the predominant field of use. The techniques predominantly employed in sentiment analysis involve traditional machine learning approaches, particularly Support Vector Machines and Naive Bayes.

Moreover, fuzzy logic, with its ability to model imprecise and uncertain information, provides a robust framework for identifying the sentiment of reviews and evaluating the potential bias that might be present in the expressions.

Although bias in sentiment analysis has attracted the attention of many researchers [25–31], sentiment analysis systems confront several challenges [32]. The effectiveness of sentiment analysis methods depends on the bias embedded in documents, such as gender or nationality bias, as well as on how well the method addresses the so-called domain adaptation problem [33].

Research studies suggest there is an urgent need to develop sentiment analysis techniques that can identify and quantify bias [29,32,34]. Bias in reviews can be attributed, among other criteria, to personal characteristics such as gender or cultural differences [27,29,35,36]. However, a few studies focus on understanding the role of cultural differences in user content generation [35–37]. This study proposes an FSE approach to calculate reviews' usefulness by incorporating bias which is embedded in reviews created by users of different nationalities. This study considers reviews' main-body and title sentiment, and articulacy as determinants of usefulness. Although fuzzy logic has been used in sentiment analysis, there is little to no research that utilizes fuzzy logic to model and analyze usefulness and bias in sentiments.

The remainder of this paper is organized as follows: 1. Literature review on reviews' usefulness and reviewers' bias; 2. Materials and Methods; 3. Results; 4. Discussion; 5. Conclusions.

1.2. Reviews' Usefulness and Reviewers' Bias in Sentiment Analysis

A plethora of reviews has flooded online platforms such as Tripadvisor, Booking, etc., since reviews have been recognized as an important source of information for customers [38,39]. As a result, when users need to focus on the most useful opinions, they encounter vast numbers of reviews that imply high search costs and information overload [40,41]. It is argued [42,43] that the adoption of helpfulness voting systems can benefit both consumers and businesses. Customers may find assistance to tackle the sheer numbers of reviews by focusing on the most appropriate reviews and businesses are thereby expected to develop revenue streams. Thus, a major research question arises with respect to how consumers identify the useful reviews and how they perceive the high-quality ones [38,39]. In the relevant literature, the usefulness of reviews has been examined by two perspectives, namely the review itself and the reviewer-related factors [38]. Review-related factors include length of review, sentiment extremity, novelty, depth, rating, and information inconsistency [38–41,44–50]. Reviewer-related factors include expertise, experience, identity, rank, and reputation [46,51]. It is also argued that since users read lots of reviews, the usefulness of a review does not solely rely on the reviews' characteristics, but also on the characteristics of the reviews that have been read previously by the user as well as on the products' context factors, such as product satisfaction, product popularity, intangibility, etc. [38]. However, research studies reach contradicting results [38]. Some studies argue that consumers prefer reviews with depth, i.e., more words in a review [42,52], while other studies argue that there is no significant relationship between review depth and usefulness [46,53]. It is argued that if the length of a review exceeds a certain threshold, then its readability diminishes as the consumer would require more time to read it [53]. However, is such a threshold the same for all reviewers? In the same vein, some studies indicate that reviews' sentiment extremity has a positive effect on usefulness [41,50], others report a negative effect [44,50], while other results show no effect at all [43,52,53]. The question that arises is whether the same extremes are perceived similarly by all reviewers. Of course, not all users express themselves or perceive review extremes the same way. Therefore,

despite the undeniable value of reviews and sentiment analysis applications, biases in users' reviews, which are often overlooked, may subsequently result in misleading and often contradicting judgements. Indeed, several research studies have focused on investigating bias in sentiment analysis [30,31]. The association of certain words and expressions with males or females as an indicator of gender bias is examined in [29]. Results indicate that women tend to use more direct language to express either positive or negative sentiments than males. Research findings suggest that there are gender differences regarding the extent to which sentiments are expressed in user reviews [25–28]. Gender bias is also found in political writing [34]. Sentiment analysis results indicate that sentiment is less positive when a political article refers to a female figure than a male. Behavioral differences were also discussed in [54], where tweets were analyzed and a method for gender identification was proposed.

Differences regarding sentiments can also be attributed to different nationalities. Users' cultural background implies that they may have different priorities, or they may seek information from different sources. Customers from Greece and Portugal prefer to rely on word of mouth rather than commercial marketing sources which are the choice of the customers from the UK and USA [36]. Users from the USA tend to appreciate reviews' evaluations from their compatriots more than those from other countries [35]. Asian hotel visitors exhibit different complaint behavior, since they seem to rather refrain from expressing their complaints publicly compared to the non-Asian visitors [37]. A research study [35] indicates that guests from western countries usually express their feelings in a more positive and informative way than others. Visitors' nationality is also identified as a discriminator factor since statistically significant differences were found in sentiments expressed by people from America, Europe, the Middle East, and Asia [55].

2. Materials and Methods

2.1. Materials and Methodology

This study proposes a methodology to assess reviews' usefulness and increase reviews' readability. It utilizes fuzzy logic in order to incorporate cultural differences and biases embedded in user reviews, and to recommend the most appropriate reviews to users according to their preferences. This study approaches cultural differences in terms of reviews' sentiments and articulacy since relevant research has identified both as determinants of reviews' usefulness [47,49]. Thus, it has the following aims:

- Analyze reviews and investigate how lenient users of different nationalities are when expressing their sentiments in their reviews, by comparing and contrasting sentiment polarity and strength in review titles and full bodies.
- Examine how informative the users from different nationalities are, i.e., examine cultural differences in terms of reviews' articulacy.
- Propose a fuzzy synthetic evaluation approach to calculate reviews' usefulness by taking into consideration users' cultural differences. Although this study considers the sentiment and articulacy determinants of reviews' usefulness, the proposed approach allows users to specify their personalized perspective of usefulness by incorporating additional features that reflect their individual biases and preferences. Figure 1 illustrates the steps of the proposed methodology.

This study collected hotel reviews, written in English, from Tripadvisor over the years 2020–2022 using the open-source web crawler Python Scrapy. Data included the review's text, its title, and the nationality of the reviewer. In total, 95,678 reviews were selected for five nationalities, namely British, American, Australian, Greek, and Dutch. These nationalities were selected since their reviews in English are available in large numbers. Reviews from any nationality would suit the purposes of this research with no difference, provided that sentiment analysis tools are available for the review language. The sample consisted of 42,678 reviews from British citizens (44.6%), 40,311 from Americans (42.1%), 9293 from Australians (9.7%), 1734 from Dutch citizens (1.8%), and 1662 from Greek citizens (1.6%). The sample is clearly biased, since the majority of the reviews selected were supplied

by reviewers whose mother tongue is English. The KNIME visual programming software platform (https://www.knime.com/) was used for data preparation and analysis. KNIME is an open-source platform that provides tools to manipulate and prepare data and machine learning algorithms for data analysis. A node is the fundamental unit in KNIME, which performs tasks such as delete stop words, create a bag of words, calculate TF-IDF scores, train a neural network, etc. Nodes are combined through drag-and-drop to develop a workflow in KNIME. KNIME has been used in several studies such as in [32,56]. The dataset collected in this research was initially cleaned, anonymized, and pre-processed in order to be imported into KNIME as a CSV file. Subsequently, the documents were checked for spelling errors, they were converted into lower cases, stop words were removed, and reviews' texts were tokenized.
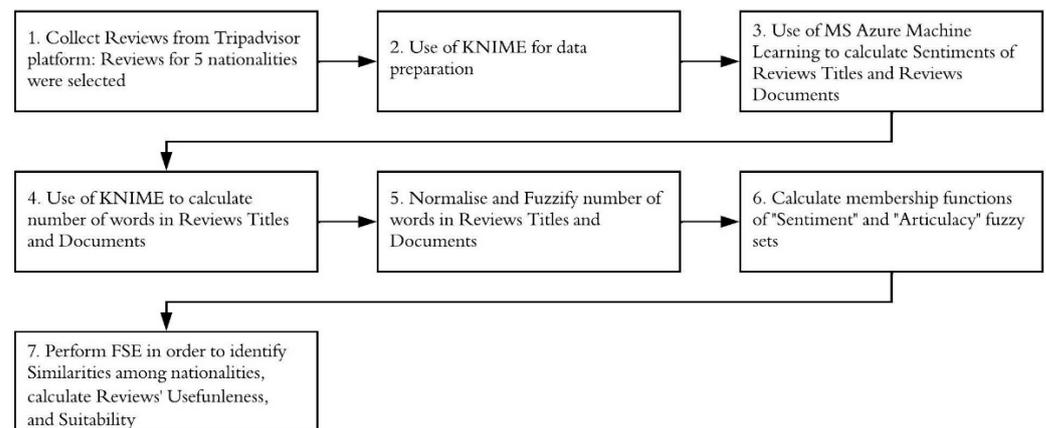

Figure 1. The steps of the proposed methodology.

This study uses MS Azure Machine Learning to calculate the sentiment strength and polarity of both the review's main body and its title. The sentiment strength returns values in the interval [0, 1]. Values close to 0 indicate negative sentiment, while values close to 1 indicate positive sentiment. The polarity is quantified with a positive, neutral, or negative value. The same MS Azure sentiment analysis service was used in [57] to calculate the sentiment of the mood that prevails in forum discussions regarding listed companies. Subsequently, it included the results in stock market data analysis. Jiang et al. (2022) [58] assessed the effectiveness of the MS Azure sentiment analysis service in conjunction with other sentiment analysis tools via metaphoric testing. In another study [59], MS Azure sentiment machine learning was used to calculate the sentiment and satisfaction expressed by patients regarding online doctor services.

2.2. Methods

This study represents sentiment and articulacy as a triangular fuzzy number (TFN). TFNs are represented by a triple (a, b, c). The membership function $f_A(x)$ of TFN $\widetilde{A}$ can be calculated according to the following equation [60]:

$$f_A(x) = \begin{cases} \frac{x-c}{a-c} & , \quad c \leq x \leq a \\[2mm] \frac{b-x}{b-a} & , \quad a \leq x \leq b \\[2mm] 0 & , \quad otherwise \end{cases} \tag{1}$$

where a, b, c are real numbers.

2.2.1. Fuzzy Relations

Fuzzy relations are important because they can describe the strength of interactions between variables [61,62]. Fuzzy relations, which are fuzzy sets, are fuzzy subsets of $X \times Y$, mapping from $X \to Y$. Let $X, Y \subseteq R$ be universal sets. Then,

$$\widetilde{R} = \{((x,y), \mu_R(x,y)) | (x,y) \in X \times Y\} \tag{2}$$

is called a fuzzy relation on $X \times Y$ [62]. A fuzzy relation on a single universe X is also a relation from X to X. It is a fuzzy tolerance relation if the two following properties define it:

Reflexivity: $\mu_R(x_i, x_i) = 1$

Symmetry: $\mu_R(x_i, x_j) = \mu_R(x_j, x_i)$

The resulting fuzzy relation is the tolerance matrix, which indicates the similarity degrees between related concepts.

If we assume fuzzy set $\widetilde{A}$ on universe X and $\widetilde{B}$ on universe Y, then the Cartesian product will result in relation R, which is contained in the Cartesian product space so that $A \times B = R \subset X \times Y$. The membership function of fuzzy relation R is calculated according to Equation (3):

$$\mu_R(x,y) = \mu_{A \times B}(x,y) = min(\mu_A(x), \mu_B(y)) \tag{3}$$

Fuzzy relations are defined in this study, in the context of the FSE, in order to represent the relationship between nationalities and the sentiments which are expressed in their reviews, as well as between nationalities and their reviews' articulacy. The membership degrees of the fuzzy relations indicate the extent to which users of a certain nationality express themselves with negative, neutral, or positive sentiments.

2.2.2. Fuzzy Synthetic Evaluation

The Fuzzy Synthetic Evaluation (FSE) has been widely used to assess multi-criteria problems [63–66]. The FSE conceptualizes a decision-making problem at three levels: the criteria, the indicators, and the alternatives [63–65]. It associates the three levels drawing on fuzzy relations. At the first level of the problem conceptualization, the FSE defines one or more criteria and their assessment grades upon which the decision-making problem is assessed. At the second level, the FSE specifies the indicators and their assessment grades. The indicators are used to measure each of the criteria. Finally, at the third level, the FSE identifies the alternative solutions to the problem. The alternatives are assessed in terms of the criteria and the corresponding indicators. The steps of FSE are as follows:

(i)   Assume that $C = \{C_i, \ i = 1, \ \dots, \ n\}$ is the set of criteria, and C_i indicates criterion (i). This study assumes the criterion "usefulness", thus, $n = 1$.

(ii) Assume that $I = \{I_j, \ j = 1, \ \dots, \ m\}$ is the set of indicators, where I_j indicates indicator (j). It consists of the "title-sentiment (t_s)", "review-sentiment (r_s)", "title-articulacy (t_a)", and the "review-articulacy (r_a)" indicators, thus, $m = 4$.

(iii) Assume that $AG = \{AG_p, \ p = 1, \ \dots, \ s\}$ is the set of assessment grades for criteria, indicators, and alternatives, with AG_p indicating assessment grades.

More specifically,

the set of assessment grades AG_p^{Ci} used for the criterion C_i = "usefulness" is defined as follows:

$$AG^{C_i} = \{AG_p^{usefulness}, \ p = 1, \ \dots, \ k_{usefulness}\} = \{Low, \ Medium, \ High\}$$

Respectively, for the indicators,

$AG^{Ij} = \{AG_p^{Ij}, \ p = 1, \ \dots, \ k_{Ij}\}$, for I_1, I_2, I_3, I_4, which in our case are $I_{ts}, I_{rs}, I_{ta}, I_{ra}$

$$AG^{I_1} = \{AG_p^{I_1}, \text{p} = 1, \ldots, k_{I_1}\} = \{AG_p^{t_s}, \text{p} = 1, \ldots, 3\} = \{\text{Negative, Neutral, Positive}\}$$
$$AG^{I_2} = \{AG_p^{I_2}, \text{p} = 1, \ldots, k_{I_2}\} = \{AG_p^{r_s}, \text{p} = 1, \ldots, 3\} = \{\text{Negative, Neutral, Positive}\}$$
$$AG^{I_3} = \{AG_p^{I_3}, \text{p} = 1, \ldots, k_{I_3}\} = \{AG_p^{t_a}, \text{p} = 1, \ldots, 3\} = \{\text{Low, Medium, High}\}$$
$$AG^{I_4} = \{AG_p^{I_4}, \text{p} = 1, \ldots, k_{I_4}\} = \{AG_p^{r_a}, \text{p} = 1, \ldots, 3\} = \{\text{Low, Medium, High}\}$$

Therefore, there are three assessment grades, $k_{C_1} = 3$, for criterion C_i = "usefulness" and three assessment grades, $k_{I_1} = k_{I_2} = k_{I_3} = k_{I_4} = 3$, for each indicator.

(iv) Assume that A = $\{r_v, \text{v} = 1, \ldots, z\}$ is the set of the alternatives, where (z) is the number of reviews that are potentially considered by the users when seeking advice for a destination.

(v) Establish the membership function matrix of fuzzy relation R for each nationality Nat = $\{\text{Nat}_g, \text{g} = 1, \ldots, d\}$,

(in our case $d = 5$, {British, US, Australian, Greek, Dutch}):

$$^{\text{Nat}_g}R = (r_{z,I_j}) = $$

		$AG_1{}^{I_1}, AG_1{}^{I_2}$ *Negative* $AG_1{}^{I_3}, AG_1{}^{I_4}$ *Low*	$AG_2{}^{I_1}, AG_2{}^{I_2}$ *Neutral* $AG_2{}^{I_3}, AG_2{}^{I_4}$ *Medium*	$AG_3{}^{I_1}, AG_3{}^{I_2}$ *Positive* $AG_3{}^{I_3}, AG_3{}^{I_4}$ *High*
for t_s and r_s				
for t_a and r_a				
$I_1,$	t_s	$r_{I_1}^{AG_1{}^{I_1}}$	$r_{I_1}^{AG_2{}^{I_1}}$	$r_{I_1}^{AG_3{}^{I_1}}$
$I_2,$	r_s	$r_{I_2}^{AG_1{}^{I_2}}$	$r_{I_2}^{AG_2{}^{I_2}}$	$r_{I_2}^{AG_3{}^{I_2}}$
$I_3,$	t_a	$r_{I_3}^{AG_1{}^{I_3}}$	$r_{I_3}^{AG_2{}^{I_3}}$	$r_{I_3}^{AG_3{}^{I_3}}$
$I_4,$	r_a	$r_{I_4}^{AG_1{}^{I_4}}$	$r_{I_4}^{AG_2{}^{I_4}}$	$r_{I_4}^{AG_3{}^{I_4}}$

$$= \qquad\qquad (4)$$

		$AG_1{}^{I_1}, AG_1{}^{I_2}$ *Negative* $AG_1{}^{I_3}, AG_1{}^{I_4}$ *Low*	$AG_2{}^{I_1}, AG_2{}^{I_2}$ *Neutral* $AG_2{}^{I_3}, AG_2{}^{I_4}$ *Medium*	$AG_3{}^{I_1}, AG_3{}^{I_2}$ *Positive* $AG_3{}^{I_3}, AG_3{}^{I_4}$ *High*
for t_s and r_s				
for t_a and r_a				
I_1	t_s	$r_{t_s}^{\text{Negative}}$	$r_{t_s}^{\text{Neutral}}$	$r_{t_s}^{\text{Positive}}$
I_2	r_s	$r_{r_s}^{\text{Negative}}$	$r_{r_s}^{\text{Neutral}}$	$r_{r_s}^{\text{Positive}}$
I_3	t_a	$r_{t_a}^{\text{Low}}$	$r_{t_a}^{\text{Medium}}$	$r_{t_a}^{\text{High}}$
I_4	r_a	$r_{r_a}^{\text{Low}}$	$r_{r_a}^{\text{Medium}}$	$r_{r_a}^{\text{High}}$

where r_{z,I_j} indicates the membership degrees to which I_j satisfies assessment grade $AG_{k_{I_j}}^{I_j}$, to the total of reviews (z).

In the R fuzzy relation matrix above, the $r_{t_s}^{\text{Negative}}$, see element (1,1), indicates the membership degree to which indicator I_1, i.e., "title sentiment—t_s", satisfies assessment grade, $AG_1^{I_1}$, i.e., "Negative". In order to calculate the $r_{t_s}^{\text{Negative}}$, we calculate the percentage of the total reviews that are designated as "Negative" [63,64,66,67]. Respectively, we calculate the percentage of the "Neutral" and "Positive" reviews.

For example, if we assume that 17% of the British reviews are rated as "Negative", 23% as "Neutral", and 60% as "Positive", then the membership degree of $r_{t_s}^{\text{Negative}}$ = 0.17. Subsequently, the membership function of the "title-sentiment—t_s", is given by (5):

$$\frac{0.17}{\text{Negative}}, \frac{0.23}{\text{Neutral}}, \frac{0.60}{\text{Positive}} \qquad (5)$$

Each of the three assessment grades is assigned to a rating factor $S_p = 1,2,3$, e.g., Negative = 1, Neutral = 2, and Positive = 3, and Low = 1, Medium = 2, and High = 3, as used in other studies [63,64,66,67], thus:

$$\frac{0.17}{\text{Negative}}, \frac{0.23}{\text{Neutral}}, \frac{0.60}{\text{Positive}} \rightarrow \frac{0.17}{1} + \frac{0.23}{2} + \frac{0.60}{3}$$

(vi) Calculate the weights WI_j for each indicator I_j. This study adopts the ordered weight averaging aggregation (OWA), which is often used in fuzzy logic [63,67]. The weights are calculated using Equations (6) and (7):

$$^{\mathrm{Nat}_g}\mathrm{Imp}_{I_j} = \sum_p \left(S_p \times r_{z,I_j} \right) \tag{6}$$

$$^{\mathrm{Nat}_g}\mathrm{WI}_{I_j} = \frac{\left(^{\mathrm{Nat}_g}\mathrm{Imp}_{I_j} \right)}{\sum_{j=1}^{m} \left(^{\mathrm{Nat}_g}\mathrm{Imp}_{I_j} \right)} \tag{7}$$

where

$^{\mathrm{Nat}_g}\mathrm{Imp}_{I_j}$ is the aggregated importance vector for indicator I_j,

S_p is the rating factor given to assessment grade $\mathrm{AG}_{k_{Ij}}^{Ij}$, and

m is the number of indicators under one criterion.
Therefore, the vector of weights for the (m) indicators is given by:

$$^{\mathrm{Nat}_g}\mathrm{WI} = \begin{bmatrix} ^{\mathrm{Nat}_g}\mathrm{WI}_{I_1} \\ ^{\mathrm{Nat}_g}\mathrm{WI}_{I_2} \\ \cdots \\ ^{\mathrm{Nat}_g}\mathrm{WI}_{I_m} \end{bmatrix} = \begin{bmatrix} ^{\mathrm{Nat}_g}\mathrm{WI}_{t_s} \\ ^{\mathrm{Nat}_g}\mathrm{WI}_{r_s} \\ ^{\mathrm{Nat}_g}\mathrm{WI}_{t_a} \\ ^{\mathrm{Nat}_g}\mathrm{WI}_{r_a} \end{bmatrix} \tag{8}$$

(vii) Calculate the weights WC_i for each criterion C_i. The weights are calculated using Equation (9):

$$^{\mathrm{Nat}_g}\mathrm{WC_i} = \frac{\left(\sum\limits_{j=1}^{m} {}^{\mathrm{Nat}_g}\mathrm{WI}_{I_j} \right)_i}{\left(\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} {}^{\mathrm{Nat}_g}\mathrm{WI}_{I_j} \right)_i} \tag{9}$$

for all indicators under criterion C_i.
This study assumes one criterion, i.e., the "usefulness" of reviews.

(viii) Establish the membership function matrix APF of the alternatives' performance for each nationality as follows:

$$\mathrm{Nat} = \{ \mathrm{Nat}_g, \, g = 1, \, \ldots, \, d \},$$

(in our case $d = 5$, {British, US, Australian, Greek, Dutch}):

$$^{\mathrm{Nat}_1}\mathrm{APF}_{r_1} = \begin{bmatrix} {}^{ts}a_{r_1,Negative} & {}^{ts}a_{r_1,Neutral} & {}^{ts}a_{r_1,Positive} \\ {}^{rs}a_{r_1,Negative} & {}^{rs}a_{r_1,Neutral} & {}^{rs}a_{r_1,Positive} \\ {}^{ta}a_{r_1,Low} & {}^{ta}a_{r_1,Medium} & {}^{ta}a_{r_1,High} \\ {}^{ra}a_{r_1,Low} & {}^{ra}a_{r_1,Medium} & {}^{ra}a_{r_1,High} \end{bmatrix}$$

$$^{\mathrm{Nat}_1}\mathrm{APF}_{r_2} = \begin{bmatrix} {}^{ts}a_{r_2,Negative} & {}^{ts}a_{r_2,Neutral} & {}^{ts}a_{r_2,Positive} \\ {}^{rs}a_{r_2,Negative} & {}^{rs}a_{r_2,Neutral} & {}^{rs}a_{r_2,Positive} \\ {}^{ta}a_{r_2,Low} & {}^{ta}a_{r_2,Medium} & {}^{ta}a_{r_2,High} \\ {}^{ra}a_{r_2,Low} & {}^{ra}a_{r_2,Medium} & {}^{ra}a_{r_2,High} \end{bmatrix} \tag{10}$$

$$\cdots$$

$$^{\mathrm{Nat}_1}\mathrm{APF}_{r_{z_{\mathrm{Nat1}}}} = \begin{bmatrix} {}^{ts}a_{r_{z_{\mathrm{Nat1}}},Negative} & {}^{ts}a_{r_{z_{\mathrm{Nat1}}},Neutral} & {}^{ts}a_{r_{z_{\mathrm{Nat1}}},Positive} \\ {}^{rs}a_{r_{z_{\mathrm{Nat1}}},Negative} & {}^{rs}a_{r_{z_{\mathrm{Nat1}}},Neutral} & {}^{rs}a_{r_{z_{\mathrm{Nat1}}},Positive} \\ {}^{ta}a_{r_{z_{\mathrm{Nat1}}},Low} & {}^{ta}a_{r_{z_{\mathrm{Nat1}}},Medium} & {}^{ta}a_{r_{z_{\mathrm{Nat1}}},High} \\ {}^{ra}a_{r_{z_{\mathrm{Nat1}}},Low} & {}^{ra}a_{r_{z_{\mathrm{Nat1}}},Medium} & {}^{ra}a_{r_{z_{\mathrm{Nat1}}},High} \end{bmatrix}$$

where $a_{z,p}$ indicates the membership degrees to which alternative z satisfies assessment grade $AG^{Ij}_{k_{Ij}}$. This study considers the set of reviews that a user may consider as the set of alternatives A for each nationality. Thus, recalling from step (iv),

$$A_{Nat_1} = \{r_{v_{Nat_1}}, \; v_{Nat_1} = 1, \ldots, z_{Nat_1}\},$$
$$A_{Nat_2} = \{r_{v_{Nat_2}}, \; v_{Nat_2} = 1, \ldots, z_{Nat_2}\},$$
$$\ldots$$
$$A_{Nat_5} = \{r_{v_{Nat_5}}, \; v_{Nat_5} = 1, \ldots, z_{Nat_5}\},$$

where

z_{Nat_1}, z_{Nat_2}, z_{Nat_3}, z_{Nat_4}, z_{Nat_5} are the total numbers of the British, US, Australian, Greek, and Dutch users' total reviews, respectively,

$$A = \sum_{h=1}^{d} A_{Nat_h}$$

and

$$z = \sum_{h=1}^{d} z_{Nat_h}.$$

(ix) Aggregate performance evaluation for alternative z using fuzzy relations [61,63,65] as shown in Equation (11):

$$\omega_z = W \otimes APF_z \tag{11}$$

The three scalars of ω_p^z represent the membership degrees of each assessment grade p, for the alternative z, thus:

$$\omega_z \rightarrow \frac{\omega_1^z}{\text{Negative}} + \frac{\omega_2^z}{\text{Neutral}} + \frac{\omega_3^z}{\text{Positive}}$$

A crisp value for ω_z can be obtained after defuzzification. This study adopts the Equations (12) and (13) used in [63] in order to calculate the score for alternative z:

$$\omega_z = 5\omega_{\text{Negative}}^z + 50\omega_{\text{Neutral}}^z + 100\omega_{\text{Positive}}^z \tag{12}$$

A final usefulness score is given by

$$\Delta_z = 100 - \omega_z \tag{13}$$

3. Results

3.1. Reviews' Sentiments Membership Functions

Our results indicate that the British exhibit more consistent behavior than the other nationalities in the sample, with respect to the sentiments expressed in their reviews' titles and reviews' full documents in all three sentiment categories (Table 1).

The sentiments the British users expressed in either their review's titles or the reviews' full documents are almost identical. All nationalities in the sample are unanimous in expressing more positive than negative sentiments to a large extent, which is a good sign for the quality of the services reviewers received during their visits.

Table 1. Title and main review sentiment percentages for each nationality in the sample.

	Sentiment					
	Title	**Review**	**Title**	**Review**	**Title**	**Review**
	Negative		**Neutral**		**Positive**	
British	4.08	4.68	6.83	6.65	89.08	88.67
American	4.73	32.21	5.89	7.10	89.37	60.67
Australian	2.60	28.74	6.46	6.40	90.94	64.86
Greek	3.55	6.26	6.68	24.13	89.77	69.61
Dutch	3.86	30.68	6.69	4.56	89.45	64.76

However, differences exist between titles' and reviews' sentiments. Differences of more positive evaluations in titles than in full texts range from a minimum of 20.16% in the Greek sample, to a maximum of 28.7% in the US sample. Similarly, more negative evaluations are found in full documents than in titles. Sentiment frequencies vary from a minimum of 2.71% more negatives in the full documents in the Greek sample, to the maximum of 27.48% in the US sample. This implies that when Greek users express negative sentiments in their titles, they do so more concisely and more accurately than the other three nationalities. The Australian and Dutch samples show larger percentage differences, which are 26.14% and 26.82%, respectively. With respect to neutral sentiments, percentage differences between titles and full reviews are rather small, ranging from almost 0 to 2.13%, with the exception of the Greek sample (17.45%). It should be remembered that people of different origin often obtain another nationality. In such cases, they shape and represent the profile of their new nationality group, incorporating their sentiments and reviews' length with those of the rest of the users. Nevertheless, reviews published on Tripadvisor do not contain information about the origin of reviewers, and the availability of such data would not affect the applicability of the proposed methodology.

To accommodate the differences in users' behavior, this study proposes to represent sentiments from different nationalities as triangular fuzzy sets $\widetilde{A}(l, m_i, u_i)$. According to [64,66,67], the membership function of each sentiment fuzzy set is formed as follows: In the sample, of the total 42,678 British reviews, 1742 expressed negative sentiment in their title, which means 4% negative reviews. Similarly, neutral titles account for 6%, and positive ones for 89%. Thus, for the British title sentiments the membership function is: $\frac{0.04}{Negative} + \frac{0.06}{Neutral} + \frac{0.89}{Positive} \rightarrow \frac{0.04}{1} + \frac{0.06}{2} + \frac{0.89}{3}$. The rating given to each assessment grade (i.e., Negative = 1, Neutral = 2, Positive = 3) is adopted by [63,64,66,67]. The membership functions for both title sentiments and main review sentiments, for each nationality, are shown in Table 2.

Table 2. The sentiment fuzzy set for both the review's title and main document for each nationality.

Nationality	Title Sentiment Fuzzy Set	Reviews' Sentiment Fuzzy Set
British	$B\widetilde{R}_S^{Title}(0.04/0.06/0.89)$	$B\widetilde{R}_S^{Review}(0.04/0.06/0.88)$
American	$US\widetilde{A}_S^{Title}(0.04/0.05/0.89)$	$US\widetilde{A}_S^{Review}(0.32/0.07/0.60)$
Australian	$A\widetilde{U}_S^{Title}(0.02/0.06/0.90)$	$A\widetilde{U}_S^{Review}(0.28/0.06/0.64)$
Greek	$G\widetilde{R}_S^{Title}(0.03/0.06/0.89)$	$G\widetilde{R}_S^{Review}(0.06/0.24/0.69)$
Dutch	$D\widetilde{U}_S^{Title}(0.03/0.06/0.89)$	$D\widetilde{U}_S^{Review}(0.30/0.04/0.64)$

Figure 2 clearly depicts the differences between the British and the US reviews' sentiment fuzzy sets, which imply the behavioral differences between the two nationalities.

The sentiment fuzzy sets with their membership functions show the level of "positiveness" in reviews for each nationality. The diagrams show that British and Greek reviewers

are inclined towards more positive comments than the other nationalities in the sample. The US, Australians and Dutch users' sentiments are closer for both titles and main reviews.
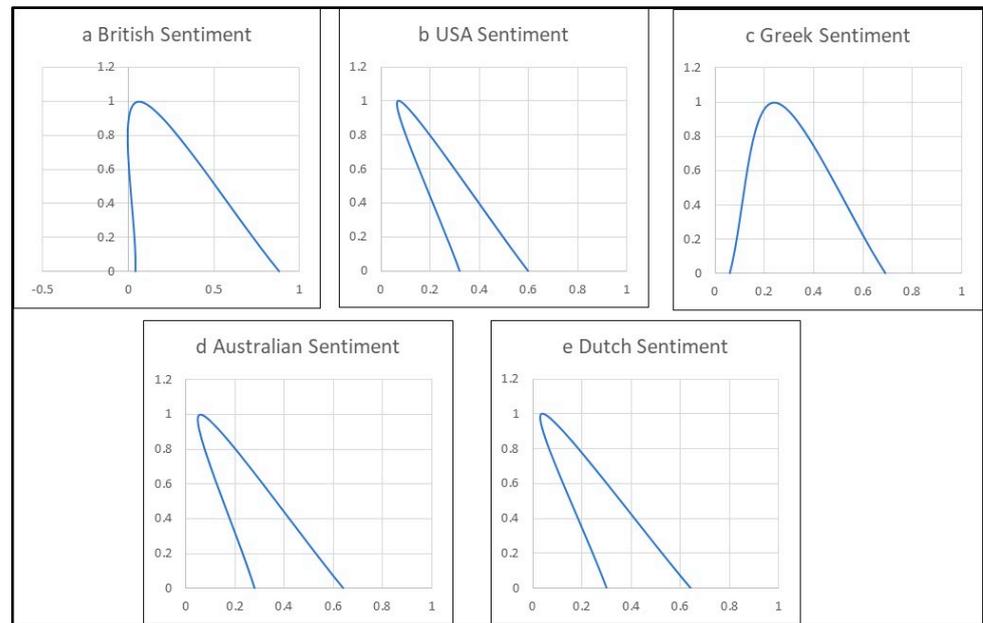

Figure 2. The reviews' sentiment fuzzy set membership function diagrams depict the behavioral differences: (**a**) British users; (**b**) US users; (**c**) Greek users; (**d**) Australian users; (**e**) Dutch users.

3.2. Reviews' Articulacy Membership Functions

Table 3 shows the results regarding the articulacy of reviews by calculating the mean and standard deviation of both the titles and the full review for each nationality.

Table 3. Average number and standard deviation of reviews' "title articulacy" and "review articulacy".

	Articulacy			
	Title		Review	
	Average Number	**Standard Deviation**	**Average Number**	**Standard Deviation**
British	4.28	2.64	97.07	41.38
American	4.59	2.62	98.60	39.90
Australian	4.37	2.57	97.23	39.46
Greek	4.34	2.79	79.76	39.60
Dutch	4.57	2.64	40.78	40.78

The normalization of the number of words in titles and main reviews is performed by applying Equation (14),

$$w_{i,j}^{norm} = \frac{w_{i,j}}{max(\sum_{i,j} w_{i,j})} \tag{14}$$

where

$w_{i,j}^{norm}$ indicates the normalized values of the number of words in the title for nationality (i) and review (j), and
$w_{i,j}$ represents the original number of words in the title for nationality and review.

Next, the fuzzification of title and review word counts is performed by using Equation (1) and the TFN membership functions shown in Table 4.

Table 4. The TFNs used to fuzzify the title and review sentiment scores and the normalized articulacy.

Linguistic Scale	Triangular Fuzzy Scale		
Negative/Low	0.00	0.00	0.25
Neutral/Medium	0.25	0.50	0.75
Positive/High	0.50	0.75	1.00

Following the fuzzification, the articulacy for both titles and reviews is characterized in terms of the fuzzy sets {Low, Medium, High}, while the sentiment is calculated as {Negative, Neutral, Positive}. The membership functions of articulacy are calculated according to [64,66,67], and this study also calculates sentiment membership functions in the same way. Of the total 42,678 British reviews in the sample, 31% of the titles were characterized as being of low, 41% as medium, and 26% as high length, respectively. Thus, for the British title articulacy set, the membership function is:

$$\frac{0.32}{Negative} + \frac{0.42}{Neutral} + \frac{0.26}{Positive} \rightarrow \frac{0.32}{1} + \frac{0.42}{2} + \frac{0.26}{3}$$

The rating given to each assessment grade (i.e., Negative = 1, Neutral = 2, Positive = 3) is adopted by [63,64,66,67]. The resulting membership functions for titles' and reviews' articulacy for each nationality are shown in Table 5.

Table 5. The fuzzy sets used to model articulacy.

Nationality	Title Articulacy Fuzzy Set	Reviews' Articulacy Fuzzy Set
British	$B\widetilde{R}_L^{Title}(0.32/0.42/0.26)$	$B\widetilde{R}_L^{Review}(0.28/0.32/0.40)$
American	$US\widetilde{A}_L^{Title}(0.25/0.45/0.30)$	$US\widetilde{A}_L^{Review}(0.25/0.33/0.42)$
Greek	$G\widetilde{R}_L^{Title}(0.30/0.44/0.26)$	$G\widetilde{R}_L^{Review}(0.44/0.31/0.25)$
Australian	$A\widetilde{U}_L^{Title}(0.28/0.45/0.27)$	$A\widetilde{U}_L^{Review}(0.26/0.35/0.39)$
Dutch	$D\widetilde{U}_L^{Title}(0.27/0.43/0.30)$	$D\widetilde{U}_L^{Review}(0.40/0.33/0.27)$

The membership functions of the articulacy fuzzy sets, shown diagrammatically in Figure 3, show the level of "length" in reviews for each nationality. They provide an indication of how informative the reviewers are.
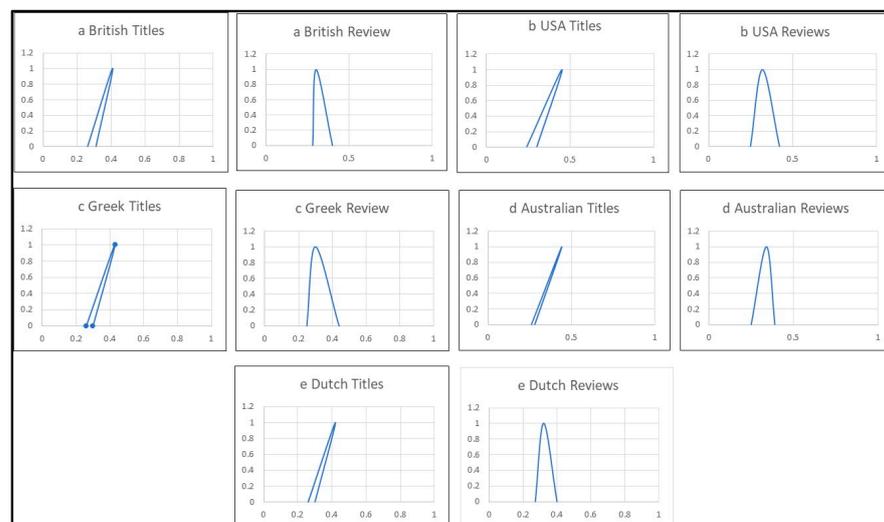
Figure 3. Fuzzy membership function diagrams indicate the differences in titles' and reviews' length between nationalities: (**a**) British users; (**b**) US users; (**c**) Greek users; (**d**) Australian users; (**e**) Dutch users.

The Figure 3 diagrams show that the British use a similar number of words in their titles to the US and the Dutch users. The Greeks are similar to the Australians who write shorter review titles. With respect to the reviews, the Greek and US reviewers exhibit similar behavior, writing longer reviews than the other nationalities in the sample. The British are similar to the Dutch and the Australians in expressing themselves with shorter reviews. An analysis of similarities and differences among nationalities that would include other indicators, such as gender, age, and preferences, may be used in order to develop a more comprehensive user profile.

3.3. Assessing Usefulness of Reviews by Incorporating Users' Biases

Drawing on the membership functions shown in Tables 2 and 5, by using Equation (4), the membership function matrix R for the British users is formed as follows:

$$
^{\text{British}}R = (r_{z,I_j}) = \begin{array}{c} grades \\ t_s \\ r_s \\ t_a \\ r_a \end{array} \begin{bmatrix} \text{Negative/Low} & \text{Neutral/Medium} & \text{Positive/High} \\ 0.04 & 0.06 & 0.90 \\ 0.05 & 0.07 & 0.88 \\ 0.32 & 0.42 & 0.26 \\ 0.30 & 0.30 & 0.40 \end{bmatrix}
$$

Similarly, membership function matrices are calculated for all nationalities in the sample.

The importance matrix and the weights for each indicator are calculated using Equations (6) and (7):

$$
^{\text{British}}\text{Imp}_{I_1} = \begin{array}{c} grades \\ t_s \end{array} \begin{bmatrix} Negative & Neutral & Positive \\ 0.04 & 0.12 & 2.70 \end{bmatrix}
$$

$$
^{\text{British}}\text{Imp}_{I_2} = \begin{array}{c} grades \\ r_s \end{array} \begin{bmatrix} Negative & Neutral & Positive \\ 0.05 & 0.14 & 2.64 \end{bmatrix}
$$

$$
^{\text{British}}\text{Imp}_{I_3} = \begin{array}{c} grades \\ t_a \end{array} \begin{bmatrix} Low & Medium & High \\ 0.32 & 0.84 & 0.78 \end{bmatrix}
$$

$$
^{\text{British}}\text{Imp}_{I_4} = \begin{array}{c} grades \\ r_s \end{array} \begin{bmatrix} Low & Medium & High \\ 0.30 & 0.60 & 1.20 \end{bmatrix}
$$

Thus,

$$
^{\text{British}}\text{Imp} == \begin{array}{c} grades \\ t_s \\ r_s \\ t_a \\ r_a \end{array} \begin{bmatrix} Negative & Neutral & Positive \\ Low & Medium & High \\ 0.04 & 0.12 & 2.70 \\ 0.05 & 0.14 & 2.64 \\ 0.32 & 0.84 & 0.78 \\ 0.30 & 0.60 & 1.20 \end{bmatrix} = \begin{bmatrix} 2.86 \\ 2.83 \\ 1.94 \\ 2.10 \end{bmatrix}
$$

Finally, the weights for the British users are:

$$
^{\text{British}}\text{WI} = \begin{bmatrix} ^{\text{British}}\text{WI}_{I_1} \\ ^{\text{British}}\text{WI}_{I_2} \\ ^{\text{British}}\text{WI}_{I_3} \\ ^{\text{British}}\text{WI}_{I_4} \end{bmatrix} = \begin{bmatrix} ^{\text{British}}\text{WI}_{t_s} \\ ^{\text{British}}\text{WI}_{r_s} \\ ^{\text{British}}\text{WI}_{t_a} \\ ^{\text{British}}\text{WI}_{r_a} \end{bmatrix} = \begin{array}{c} t_s \\ r_s \\ t_a \\ r_a \end{array} \begin{bmatrix} 0.29 \\ 0.29 \\ 0.20 \\ 0.22 \end{bmatrix}
$$

Similarly, the weights for each nationality are:

$$\text{USA}\text{WI} = \begin{matrix} t_s \\ r_s \\ t_a \\ r_a \end{matrix} \begin{bmatrix} 0.30 \\ 0.24 \\ 0.22 \\ 0.24 \end{bmatrix}, \qquad \text{Greek}\text{WI} = \begin{matrix} t_s \\ r_s \\ t_a \\ r_a \end{matrix} \begin{bmatrix} 0.31 \\ 0.28 \\ 0.21 \\ 0.20 \end{bmatrix},$$

$$\text{Australian}\text{WI} = \begin{matrix} t_s \\ r_s \\ t_a \\ r_a \end{matrix} \begin{bmatrix} 0.31 \\ 0.25 \\ 0.21 \\ 0.23 \end{bmatrix}, \qquad \text{Dutch}\text{WI} = \begin{matrix} t_s \\ r_s \\ t_a \\ r_a \end{matrix} \begin{bmatrix} 0.31 \\ 0.26 \\ 0.22 \\ 0.21 \end{bmatrix}$$

The results indicate the differences as well as the similarities among the nationalities, which have been diagrammatically depicted in Figures 2 and 3. Larger differences are found in title and review sentiments as compared to the differences in articulacy.

In order to calculate usefulness of reviews, assume "Review-1" as an alternative in the FSE model. "Review-1" can be a certain review or a collection of reviews over a period of time, e.g., 5 years. The aggregation of reviews over a period of time is suggested as especially useful [38]. By using MS Azure sentiment analysis, "Review-1" sentiment polarity is quantified with a positive, neutral, or negative value. The numerical values of Review-1 articulacy for both title and review are then fuzzified using Equation (1) and the TFN membership functions shown in Table 4.

Then the performance membership function matrix for Review-1, using Equation (10), follows:

$$\text{APF}^{\text{Review}-1} = (a_{z,p}) = \begin{matrix} \text{grades} \\ t_s \\ r_s \\ t_a \\ r_a \end{matrix} \begin{bmatrix} \textit{Negative} & \textit{Neutral} & \textit{Positive} \\ \textit{Low} & \textit{Medium} & \textit{High} \\ 0.00 & 0.00 & 0.25 \\ 0.25 & 0.50 & 0.75 \\ 0.25 & 0.50 & 0.75 \\ 0.50 & 0.75 & 1.00 \end{bmatrix}$$

The aggregated performance evaluation, regarding the usefulness of alternative "Review-1", is calculated using Equation (11):

$$\omega_{\text{Review}-1}^{\text{British}} = \begin{matrix} \text{High} \\ \text{Medium} \\ \text{Low} \end{matrix} \begin{bmatrix} 0.23 \\ 0.41 \\ 0.66 \end{bmatrix}$$

Similarly, for the rest of the nationalities in the sample:

$$\omega_{\text{Review}-1}^{\text{USA}} = \begin{matrix} \text{High} \\ \text{Medium} \\ \text{Low} \end{matrix} \begin{bmatrix} 0.23 \\ 0.41 \\ 0.66 \end{bmatrix},$$

$$\omega_{\text{Review}-1}^{\text{Greek}} = \begin{matrix} \text{High} \\ \text{Medium} \\ \text{Low} \end{matrix} \begin{bmatrix} 0.22 \\ 0.39 \\ 0.64 \end{bmatrix},$$

$$\omega_{\text{Review}-1}^{\text{Australian}} = \begin{matrix} \text{High} \\ \text{Medium} \\ \text{Low} \end{matrix} \begin{bmatrix} 0.23 \\ 0.40 \\ 0.65 \end{bmatrix},$$

$$\omega_{\text{Review}-1}^{\text{Dutch}} = \begin{matrix} \text{High} \\ \text{Medium} \\ \text{Low} \end{matrix} \begin{bmatrix} 0.22 \\ 0.39 \\ 0.64 \end{bmatrix}$$

Finally, by using Equations (12) and (13), the "Usefulness Score" = 12.80. This score indicates how useful it is for the British users to read "Review-1". Similarly, the usefulness score is calculated for each nationality as follows:

"Usefulness Score (British)" = 12.80,
"Usefulness Score (USA)" = 12.89,
"Usefulness Score (Greek)" = 14.65,
"Usefulness Score (Australian)" = 13.44,
"Usefulness Score (Dutch)" = 14.73.

The "Usefulness Score" shows that "Review-1" is not equally useful for all nationalities. It is more useful for the Greeks, the Australians, and the Dutch to read, but not so for the British and the US users. The difference in usefulness is attributed to the fact that not all users exhibit the same behavior when expressing themselves. Some may express themselves by using more words in their reviews, or others may be more lenient than others, or exhibit a behavior in between. As already discussed in the relevant literature [35,37,55], not all nationalities will necessarily select the same set of reviews. The "Usefulness Score" can be used to rank reviews or a collection of reviews and assist users to focus on the most useful reviews.

4. Discussion

Behavioral differences exist among different nationalities in the way they express their sentiments. Some nationalities express their sentiments already in their reviews' titles, while others become more precise in their full reviews. Therefore, titles do not always convey in full what users' sentiments imply. Instead, they may create a false perception of other reviewers' experiences. An explanation may be that reviewers find more space in reviews' main documents to express their sentiments in more detail. However, not all nationalities exhibit the same behavior. Differences can also be found with respect to sentiments grades as well. In particular, the British, followed by the Greeks, exhibit a more consistent behavior in expressing themselves. Users of both nationalities express themselves concisely in their reviews' titles. By reading just the titles of reviews written by a British or a Greek reviewer, one can understand the reviewer's negative feelings clearly. With respect to neutral sentiments, Greek users exhibit a varying behavior, as opposed to the rest of the nationalities studied here. Thus, neutral judgments are not so informative when they are expressed by Greek users. Regarding positive sentiments, reviews tend to be more lenient in their titles and more critical in their full documents. With the exception of the British users' rather balanced evaluations, positive scores are more frequent in titles than in full review documents for all other nationalities in the sample. Therefore, when positive sentiments are expressed, it is suggested that one should read the whole review document to have a more detailed understanding of the users' experiences. Reading the review title on its own will not suffice to reveal the thoughts of the user. The results of this research are in line with other studies [35,36,55], and suggest that nationality may imply differences in the way users write reviews and express their sentiments.

This study also indicates that there are differences among nationalities in the sample, with respect to the number of words in both the titles and the main review documents. Although differences may not be as profound as in sentiments, the results show that the British, the American, the Greek, and the Australian reviewers exhibit similar behavior regarding their reviews' length. However, the Dutch reviewers are more frugal in their full reviews, since their average number of words is approximately half the average number of the rest of the nationalities.

This study suggests that fuzzy logic can be used to represent the differences in behavior among nationalities as fuzzy sets, but also to calculate the "usefulness" of a review by utilizing the FSE. Fuzzy sets provide the means for quantifying the differences among users that imply the biases of different nationalities. The FSE of reviews' usefulness can be used to suggest users read the most suitable reviews for them through fuzzy relations. The most useful reviews are those which reflect their behavior better. Therefore, the proposed approach can be used within the context of a helpfulness voting system or online review platform [38] that aims to assist users in identifying useful reviews that reflect their personalized preferences, behavior, and perception. The current work considers usefulness

as the only criterion. The FSE and fuzzy relations can be used to combine several other characteristics of the reviews, e.g., publication date, features mentioned in reviews, etc.

With respect to the limitations of this study, this research uses only one sentiment analysis method and focuses only on reviews' sentiment polarity, strength, and articulacy for certain nationalities. The plethora of data that can be collected from platforms such as Tripadvisor, Booking, etc., and the subsequent analysis of additional features such as gender, date of visit, date of review publication, etc., would result in more detailed insights of users' ways of expressing their feelings, thus providing a more comprehensive view of biases. Furthermore, testing the efficiency of the proposed approach to accurately assess users' sentiment when combined with sentiment analysis algorithms, would probably lead to developing domain-specific sentiment analysis methods.

Future research may focus on comparing the methodology presented in this paper with learning-to-rank algorithms used in information retrieval, in assessing web document quality, etc. Furthermore, research may also aim at extending the modelling approach to include a broader perspective of users' behavior. The proposed methodology can be applied to analyze reviews' usefulness from users from any nationality, age group, purpose of travel, etc. Research efforts may focus on developing models that capture the different ways that users express themselves, since not all users perceive the sentiment grades equally. Thus, the expressions "good service" or "enjoyable experience" may not convey exactly the same meaning for all users. Since fuzzy logic allows for dealing with subjectivity, future research efforts may attempt to define different fuzzy sets for the same concepts in order to reflect how different users perceive same or similar expressions.

5. Conclusions

This study suggests that since bias affects users' perception of review sentiment and length, fuzzy logic provides the necessary theoretical and methodological foundation to measure reviews' usefulness. User reviews' sentiment is subjective. Bias attributed to gender, nationality, etc., identified via sentiment analysis has been examined in many studies. Fuzzy logic provides the means to deal with impartial information which is often found in reviews and to represent the subjectivity that is embedded in how people with different cultural backgrounds, from different age groups, with varying preferences, express their experiences. This study analyzed users' reviews and calculated the membership functions of fuzzy sets that exhibit the bias, as well as the similarities, displayed among users' different nationalities. Our results suggest that sentiments and articulacy of review titles and full documents, if not overrepresented or underrated, can provide valuable insights in understanding how users are expressing themselves, thus improving the accuracy of sentiment analysis techniques and reviews' usefulness.

References

1.　Liu, B. *Sentiment Analysis Essentials Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*; Cambridge University Press: Cambridge, UK, 2015.
2.　Yu, L.; Wang, D.; Liu, D.; Liu, Y. Research on Intelligence Computing Models of Fine-Grained Opinion Mining in Online Reviews. *IEEE Access* **2019**, *7*, 116900–116910. [CrossRef]
3.　Oktaviani, V.; Warsito, B.; Yasin, H.; Santoso, R.; Suparti, S. Sentiment Analysis of e-Commerce Application in Traveloka Data Review on Google Play Site Using Naïve Bayes Classifier and Association Method 2020. *J. Phys. Conf. Ser.* **2021**, *1943*, 012147. [CrossRef]
4.　Wang, C.; Zhu, X.; Yan, L. Sentiment Analysis for E-Commerce Reviews Based on Deep Learning Hybrid Model. In Proceedings of the 5th International Conference on Signal Processing and Machine Learning (SPML), Dalian, China, 4–6 August 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 38–46.
5.　Savci, P.; Das, B. Prediction of the Customers' Interests Using Sentiment Analysis in e-Commerce Data for Comparison of Arabic, English, and Turkish Languages. *J. King Saud Univ.-Comput. Inf. Sci.* **2023**, *35*, 227–237. [CrossRef]
6.　Hossain, M.J.; Das Joy, D.; Das, S.; Mustafa, R. Sentiment Analysis on Reviews of E-Commerce Sites Using Machine Learning Algorithms. In Proceedings of the International Conference on Innovations in Science, Engineering and Technology (ICISET), Chattogram, Bangladesh, 25–28 February 2022; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2022; pp. 522–527.
7.　Rajesh, P.; Suseendran, G. Prediction of N-Gram Language Models Using Sentiment Analysis on E-Learning Reviews. In Proceedings of the International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 17–19 June 2020; pp. 510–514.
8.　dos Santos Alencar, M.A.; de Magalhães Netto, J.F.; de Morais, F. A Sentiment Analysis Framework for Virtual Learning Environment. *Appl. Artif. Intell.* **2021**, *35*, 520–536. [CrossRef]
9.　Yan, X.; Jian, F.; Sun, B. SAKG-BERT: Enabling Language Representation with Knowledge Graphs for Chinese Sentiment Analysis. *IEEE Access* **2021**, *9*, 101695–101701. [CrossRef]
10.　Sayeedunnisa, S.F.; Hijab, M. Impact of e-Learning in Education Sector: A Sentiment Analysis View. In Proceedings of the IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 21–23 December 2022; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2022; pp. 1–5.
11.　Singh, L.K.; Devi, R.R. Analysis of Student Sentiment Level Using Perceptual Neural Boltzmann Machine Learning Approach for E-learning Applications. In Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2022, Lalitpur, Nepal, 20–22 July 2022; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2022; pp. 1270–1276.
12.　Khanam, Z. Sentiment Analysis of User Reviews in an Online Learning Environment: Analyzing the Methods and Future Prospects. *Eur. J. Educ. Pedagog.* **2023**, *4*, 209–217. [CrossRef]
13.　Krouska, A.; Troussas, C.; Virvou, M. Deep Learning for Twitter Sentiment Analysis: The Effect of Pre-Trained Word Embedding. In *Machine Learning Paradigms*; Tsihrintzis, G., Jain, L., Eds.; Learning and Analytics in Intelligent Systems; Springer: Cham, Switzerland, 2020; pp. 111–124.
14.　Li, L.; Wu, Y.; Zhang, Y.; Zhao, T. Time+User Dual Attention Based Sentiment Prediction for Multiple Social Network Texts with Time Series. *IEEE Access* **2019**, *7*, 17644–17653. [CrossRef]
15.　Wang, T.; Lu, K.; Chow, K.P.; Zhu, Q. COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model. *IEEE Access* **2020**, *8*, 138162–138169. [CrossRef] [PubMed]
16.　Alattar, F.; Shaalan, K. Using Artificial Intelligence to Understand What Causes Sentiment Changes on Social Media. *IEEE Access* **2021**, *9*, 61756–61767. [CrossRef]
17.　Silva, H.; Andrade, E.; Araújo, D.; Dantas, J. Sentiment Analysis of Tweets Related to SUS before and during COVID-19 Pandemic. *IEEE Lat. Am. Trans.* **2022**, *20*, 6–13. [CrossRef]
18.　Rodríguez-Ibánez, M.; Casánez-Ventura, A.; Castejón-Mateos, F.; Cuenca-Jiménez, P.M. A review on Sentiment Analysis from Social Media Platforms. *Expert Syst. Appl.* **2023**, *223*, 119862. [CrossRef]
19.　Krouska, A.; Troussas, C.; Virvou, M. Comparative Evaluation of Algorithms for Sentiment Analysis over Social Networking Services. *J. Univers. Comput. Sci. (JUCS)* **2017**, *23*, 755–768.
20.　Usher, J.; Morales, L.; Dondio, P. BREXIT: A Granger Causality of Twitter Political Polarisation on the FTSE 100 Index and the Pound. In Proceedings of the 2nd IEEE International Conference on Artificial Intelligence and Knowledge Engineering, AIKE 2019, Cagliari, Italy, 3–5 June 2019; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019; pp. 51–54.
21.　Shaghaghi, N.; Calle, A.M.; Manuel Zuluaga Fernandez, J.; Hussain, M.; Kamdar, Y.; Ghosh, S. Twitter Sentiment Analysis and Political Approval Ratings for Situational Awareness. In Proceedings of the IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), Tallinn, Estonia, 14–22 May 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 59–65.
22.　Schmale, A.; Mittendorf, V. Detecting Negative Campaigning on Twitter Against The Greens. In Proceedings of the Ninth IEEE International Conference on Social Networks Analysis, Management and Security (SNAMS), Milan, Italy, 29 November–1 December 2022; pp. 1–8.
23.　Orellana, S.; Bisgin, H. Using Natural Language Processing to Analyze Political Party Manifestos from New Zealand. *Information* **2023**, *14*, 152. [CrossRef]

24. Ligthart, A.; Catal, C.; Tekinerdogan, B. Systematic Reviews in Sentiment Analysis: A Tertiary Study. *Artif. Intell. Rev.* **2021**, *54*, 4997–5053. [CrossRef]

25. Thelwall, M.; Wilkinson, D.; Uppal, S. Data Mining Emotion in Social Network Communication: Gender Differences in MySpace. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 190–199. [CrossRef]

26. Volkova, S.; Yoram, B. On Predicting Sociodemographic Traits and Emotions from Communications in Social Networks and their Implications to Online Self-Disclosure. *Cyberpsychology Behav. Soc. Netw.* **2015**, *12*, 726–736. [CrossRef] [PubMed]

27. Babac, M.B.; Podobnik, V. A Sentiment Analysis of Who Participates, How and Why, at Social Media Sport Websites: How Differently Men and Women Write about Football. *Online Inf. Rev.* **2016**, *40*, 814–833. [CrossRef]

28. Rangel, F.; Rosso, P. On the Impact of Emotions on Author Profiling. *Inf. Process. Manag.* **2016**, *52*, 73–92. [CrossRef]

29. Thelwall, M. Gender Bias in Sentiment Analysis. *Online Inf. Rev.* **2018**, *42*, 45–57. [CrossRef]

30. Rajshakhar, P.; Bosu, A.; Kazi, S.Z. Expressions of Sentiments during Code Reviews: Male vs. Female. In Proceedings of the 26th IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), Hangzhou, China, 24–27 February 2019; IEEE: Piscataway, NJ, USA; pp. 26–37.

31. Sun, T.; Gaut, A.; Tang, S.; Huang, Y.; ElSherief, M.; Zhao, J.; Mirza, D.; Belding, E.; Chang, K.; Wang, W.Y. Mitigating Gender Bias in Natural Language Processing: Literature Review. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 1630–1640.

32. Ordenes, V.F.; Silipo, R. Machine Learning for Marketing on the KNIME Hub: The Development of a Live Repository for Marketing Applications. *J. Bus. Res.* **2021**, *137*, 393–410. [CrossRef]

33. López, M.; Valdivia, A.; Martínez-Cámara, E.; Luzón, M.V.; Herrera, F. E2SAM: Evolutionary Ensemble of Sentiment Analysis Methods for Domain Adaptation. *Inf. Sci.* **2019**, *480*, 273–286. [CrossRef]

34. Davis, S.R.; Worsnop, C.J.; Hand, E.M. Gender Bias Recognition in Political News Articles. *Mach. Learn. Appl.* **2022**, *8*, 100304. [CrossRef]

35. Kim, J.M.; Jun, M.; Kim, C.K. The Effects of Culture on Consumers' Consumption and Generation of Online Reviews. *J. Interact. Mark.* **2018**, *43*, 134–150. [CrossRef]

36. Litvin, S.W. Hofstede, Cultural Differences, and TripAdvisor hotel Reviews. *Int. J. Tour. Res.* **2019**, *21*, 712–717. [CrossRef]

37. Ngai, E.W.T.; Heung, V.C.S.; Wong, Y.H.; Chan, F.K.Y. Consumer Complaint Behaviour of Asians and Non-Asians about Hotel Services: An Empirical Analysis. *Eur. J. Mark.* **2007**, *41*, 1375–1391. [CrossRef]

38. Choi, H.S.; Leon, S. An Empirical Investigation of Online Review Helpfulness: A Big Data Perspective. *Decis. Support Syst.* **2020**, *139*, 113403. [CrossRef]

39. Zhang, X.; Zhang, X.; Liang, S.; Yang, Y.; Law, R. Infusing New Insights: How Do Review Novelty and Inconsistency Shape the Usefulness of Online Travel Reviews. *Tour. Manag.* **2023**, *96*, 104703. [CrossRef]

40. Park, D.H.; Lee, J. eWOM Overload and its Effect on Consumer Behavioral Intention Depending on Consumer Involvement. *Electron. Commer. Res. Appl.* **2008**, *7*, 386–398. [CrossRef]

41. Siering, M.; Muntermann, J.; Rajagopalan, B. Explaining and Predicting Online Review Helpfulness: The Role of Content and Reviewer-Related signals. *Decis. Support Syst.* **2018**, *108*, 1–12. [CrossRef]

42. Hong, H.; Xu, D.; Wang, G.A.; Fan, W. Understanding the Determinants of Online Review Helpfulness: A Meta-Analytic Investigation. *Decis. Support Syst.* **2017**, *102*, 1–11. [CrossRef]

43. Lee, S.; Choeh, J.Y. The Interactive Impact of Online Word-Of-Mouth and Review Helpfulness on Box Office Revenue. *Manag. Decis.* **2018**, *56*, 849–866. [CrossRef]

44. Cao, Q.; Duan, W.; Gan, Q. Exploring Determinants of Voting for the "Helpfulness" of Online User Reviews: A Text Mining Approach. *Decis. Support Syst.* **2011**, *50*, 511–521. [CrossRef]

45. Baek, H.; Ahn, J.; Choi, Y. Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues. *Int. J. Electron. Commer.* **2012**, *17*, 99–126. [CrossRef]

46. Racherla, P.; Friske, W. Perceived "Usefulness" of Online Consumer Reviews: An Exploratory Investigation Across Three Services Categories. *Electron. Commer. Res. Appl.* **2012**, *11*, 548–559. [CrossRef]

47. Liu, Z.; Park, S. What Makes a Useful Online Review? Implication for Travel Product Websites. *Tour. Manag.* **2015**, *47*, 140–151. [CrossRef]

48. Zhang, Y.; Lin, Z. Predicting the Helpfulness of Online Product Reviews: A Multilingual Approach. *Electron. Commer. Res. Appl.* **2018**, *27*, 1–10. [CrossRef]

49. Chatterjee, S. Drivers of Helpfulness of Online Hotel Reviews: A Sentiment and Emotion Mining Approach. *Int. J. Hosp. Manag.* **2020**, *85*, 102356. [CrossRef]

50. Bigne, E.; Ruiz, C.; Cuenca, A.; Perez, C.; Garcia, A. What Drives the Helpfulness of Online Reviews? A Deep Learning Study of Sentiment Analysis, Pictorial Content and Reviewer Expertise for Mature Destinations. *J. Destin. Mark. Manag.* **2021**, *20*, 100570. [CrossRef]

51. Zhu, L.; Yin, G.; He, W. Is this Opinion Leader's Review Useful? Peripheral Cues for Online Review Helpfulness. *J. Electron. Commer. Res.* **2014**, *15*, 2014.

52. Mudambi, S.M.; Schuff, D. What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Q.* **2010**, *34*, 185–200. [CrossRef]

53. Chua, A.Y.K.; Banerjee, S. Understanding Review Helpfulness as a Function of Reviewer Reputation, Review Rating, and Review Depth. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 354–362. [CrossRef]
54. Onikoyi, B.; Nnamoko, N.; Korkontzelos, I. Gender Prediction with descriptive Textual Data Using a Machine Learning Approach. *Nat. Lang. Process. J.* **2023**, *4*, 100018. [CrossRef]
55. Rita, P.; Ramos, R.; Borges-Tiago, M.T.; Rodrigues, D. Impact of the Rating System on Sentiment and Tone of Voice: A Booking.com and TripAdvisor Comparison Study. *Int. J. Hosp. Manag.* **2022**, *104*, 103245. [CrossRef]
56. Gitto, S.; Mancuso, P. Improving Airport Services Using Sentiment Analysis of the Websites. *Tour. Manag. Perspect.* **2017**, *22*, 132–136. [CrossRef]
57. Wojarnik, G. Sentiment Analysis as a Factor Included in the Forecasts of Price Changes in the Stock Exchange. *Procedia Comput. Sci.* **2021**, 3176–3183. [CrossRef]
58. Jiang, M.; Chen, T.Y.; Wang, S. On the Effectiveness of Testing Sentiment Analysis Systems with Metamorphic Testing. *Inf. Softw. Technol.* **2022**, *150*, 106966. [CrossRef]
59. Dhakate, N.; Joshi, R. Classification of Reviews of E-Healthcare Services to Improve Patient Satisfaction: Insights from an Emerging Economy. *J. Bus. Res.* **2023**, *164*, 114–115. [CrossRef] [PubMed]
60. Lin, H.Y.; Hsu, P.Y.; Sheen, G.J. A Fuzzy-Based Decision-Making Procedure for Data Warehouse System Selection. *Expert Syst. Appl.* **2007**, *32*, 939–953. [CrossRef]
61. Ross, T.J. *Fuzzy Logic with Engineering Applications*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2010.
62. Luukka, P. Similarity Classifier Using Similarities Based on Modified Probabilistic Equivalence Relations. *Knowl. Based Syst.* **2009**, *22*, 57–62. [CrossRef]
63. Hu, G.; Liu, H.; Chen, C.; He, P.; Li, J.; Hou, H. Selection of Green Remediation Alternatives for Chemical Industrial Sites: An Integrated Life Cycle Assessment and Fuzzy Synthetic Evaluation Approach. *Sci. Total Environ.* **2022**, *845*, 157211. [CrossRef]
64. Xu, Y.; Yeung, J.F.Y.; Chan, A.P.C.; Chan, D.W.M.; Wang, S.Q.; Ke, Y. Developing a Risk Assessment Model for PPP projects in China—A Fuzzy Synthetic Evaluation Approach. *Autom. Constr.* **2010**, *19*, 929–943. [CrossRef]
65. Akter, M.; Jahan, M.; RuKabir, R.; Karim, D.S.; Haque, A.; Munsur, M.; Salehin, M. Risk Assessment Based on Fuzzy Synthetic Evaluation Method. *Sci. Total Environ.* **2019**, *658*, 818–829. [CrossRef] [PubMed]
66. Zhao, X.; Hwang, B.G.; Gao, Y. A Fuzzy Synthetic Evaluation Approach for Risk Assessment: A Case of Singapore's Green Projects. *J. Clean. Prod.* **2016**, *115*, 203–213. [CrossRef]
67. Yager, R.R. On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making. *IEEE Trans. Syst. Man. Cybern.* **1988**, *18*, 183–190. [CrossRef]